Error and Erasure Exponents for the Broadcast Channel with Degraded Message Sets

Vincent Y. F. Tan
Department of Electrical and Computer Engineering,
Department of Mathematics,
National University of Singapore
Email: vtan@nus.edu.sg

***Abstract*—Error and erasure exponents for the broadcast channel with degraded message sets are analyzed. The focus of our error probability analysis is on the main receiver where, nominally, both messages are to be decoded. A two-step decoding algorithm is proposed and analyzed. This receiver first attempts to decode both messages, failing which, it attempts to decode only the message representing the coarser information, i.e., the cloud center. This algorithm reflects the intuition that we should decode both messages only if we have confidence in the estimates; otherwise one should only decode the coarser information. The resulting error and erasure exponents, derived using the method of types, are expressed in terms of a penalized form of the modified random coding error exponent.**

***Index Terms*—Erasure decoding, Broadcast channel, Degraded message sets, Error exponents, Method of types**

I. Introduction

In 1968, Forney [1] derived exponential error bounds for decoding with an erasure option. In this seminal paper, Forney used a generalization of the Neyman-Pearson lemma to derive an optimum decoding rule for point-to-point channel coding, where the decoder is allowed to output an erasure symbol should it not be sufficiently confident to output a message. Based on this rule, Forney used Gallager-style bounding techniques to derive exponents for the undetected and total (undetected plus erasure) error probabilities.

This work led to many follow-up works. We only mention a subset of the literature here. We mainly build on the exposition in Csiszár and Körner [2, Thm. 10.11] in which universally attainable erasure and error exponents were derived. Telatar [3] also derived and analyzed a erasure decoding rule based on a general decoding metric. Moulin [4] considered a Neyman-Pearson formulation for universal erasure decoding. Merhav [5] used the type-class enumerator method to analyze the Forney decoding rule and showed that the derived exponents are at least as good as those Forney derived. This was subsequently sharpened by Somekh-Baruch and Merhav [6] who derived the *exact* random coding exponents. Sabbag and Merhav [7] analyzed the error and erasure exponents for channels with noncausal state information (Gel'fand-Pinsker coding).

However, no generalization of the study of erasure exponents to multi-user systems with multiple messages has been published.[1] In this paper, we study the broadcast channel with degraded message sets, also known as the asymmetric broadcast channel (ABC). For this channel, the main receiver desires to decode two messages M_1 and M_2 while the secondary receiver only desires to decode the private message M_2. The capacity region, derived by Körner and Marton [8] is

$$\mathscr{C}=\bigcup_{P_{UX}}\left\{(R_1,R_2)\in\mathbb{R}_+^2: \begin{array}{r} R_1\le I(X\wedge Y|U)\\ R_2\le I(U\wedge Z)\\ R_1+R_2\le I(X\wedge Y)\end{array}\right\}. \quad (1)$$

Error exponents (without erasures) were derived by Körner and Sgarro [9] and improved by Kaspi and Merhav [10]. We go beyond these analyses to derive erasure and error exponents for the ABC. The resulting exponents involve a *penalized* form of the modified random coding error exponent derived in [2, Ch. 10] and reflects the superposition coding scheme [11] used to achieve the region in (1).

II. Preliminaries and System Model

We adopt the notation from [2]. Random variables (e.g., X) and their realizations (e.g., x) are in upper- and lower-case respectively. All random variables take values on finite sets, denoted in calligraphic font (e.g., $\mathcal{X}$). For a sequence $\mathbf{x}=(x_1,\ldots,x_n)\in\mathcal{X}^n$, its *type* is the distribution $P_{\mathbf{x}}(a)=\frac{1}{n}\sum_{i=1}^n\mathbf{1}\{x_i=a\}, a\in\mathcal{X}$. The set of types with denominator n supported on alphabet $\mathcal{X}$ is denoted as $\mathscr{P}_n(\mathcal{X})$. The *type class* of P is denoted as $\mathcal{T}_P$. For $\mathbf{x}\in\mathcal{T}_P$, the set of sequences $\mathbf{y}\in\mathcal{Y}^n$ such that $(\mathbf{x},\mathbf{y})$ has joint type $P\times V$ is the V*-shell* $\mathcal{T}_V(\mathbf{x})$. Let $\mathscr{V}_n(\mathcal{Y};P)$ be the family of stochastic matrices $V:\mathcal{X}\to\mathcal{Y}$ for which the V-shell of a sequence of type $P\in\mathscr{P}_n(\mathcal{X})$ is not empty. Information-theoretic quantities are denoted in the usual way. For example $I(P,V)$ and $I_{P\times V}(X\wedge Y)$ denote the mutual information where these expressions indicate that the joint distribution of (X,Y) is $P\times V$. In addition, $\hat{I}(\mathbf{x}\wedge\mathbf{y})$ is the empirical mutual information of $(\mathbf{x},\mathbf{y})$, i.e., if $\mathbf{x}\in\mathcal{T}_P$ and $\mathbf{y}\in\mathcal{T}_V(\mathbf{x})$, then, $\hat{I}(\mathbf{x}\wedge\mathbf{y}) := I(P,V)$. We use $a_n\doteq b_n$ to mean *equality to first-order in the exponent*, i.e., $\frac{1}{n}\log\frac{a_n}{b_n}\to 0$; exponential inequalities $\dot{\le}$ and $\dot{\ge}$ are defined similarly. Finally, $|a|^+:=\max\{a,0\}$ and $[a]:=\{1,\ldots,\lceil a\rceil\}$ for any $a\in\mathbb{R}$.

A *discrete memoryless broadcast channel* $W:\mathcal{X}\to\mathcal{Y}\times\mathcal{Z}$ is a stochastic map from a finite input alphabet $\mathcal{X}$ to the Cartesian product of two finite output alphabets $\mathcal{Y}$ and $\mathcal{Z}$. An (n,R_1,R_2)*-code* is a tuple of maps $f:[2^{nR_1}]\times[2^{nR_2}]\to$

[1] Moulin mentioned in [4, Sec. VIII] that the analysis contained therein "has been extended to compound MACs" but this extension is unpublished.

$\mathcal{X}^n$ and $\varphi_1 : \mathcal{Y}^n \to ([2^{nR_1}] \cup \{\mathrm{e}_1\}) \times ([2^{nR_2}] \cup \{\mathrm{e}_2\})$ and $\varphi_2 : \mathcal{Z}^n \to [2^{nR_2}] \cup \{\mathrm{e}_2\}$, where e_j is the erasure symbol for message $M_j, j = 1, 2$ which is uniform on the *message set* $\mathcal{M}_j := [2^{nR_j}]$. Let $W_{\mathcal{Y}}$ and $W_{\mathcal{Z}}$ be the marginals of W.

We may define error probabilities for both terminals $\mathcal{Y}$ and $\mathcal{Z}$. However, we will focus exclusively on terminal $\mathcal{Y}$ as the analysis is more interesting and non-standard. We are concerned with four different error probabilities at terminal $\mathcal{Y}$. Let $\mathcal{D}_{m_1}, m_1 \in \mathcal{M}_1 := [2^{nR_1}]$ and $\mathcal{D}_{m_2}, m_2 \in \mathcal{M}_2 := [2^{nR_2}]$ be the disjoint decoding regions associated to messages m_1 and m_2 respectively. This means that $\mathcal{D}_{m_1} := \cup_{m_2 \in \mathcal{M}_2}\{\mathbf{y} : \varphi_1(\mathbf{y}) = (m_1, m_2)\}$ and similarly for $\mathcal{D}_{m_2}$. Note that because we allow erasures $\mathcal{Y}^n \backslash \cup_{m_1 \in \mathcal{M}_1} \mathcal{D}_{m_1}$ need not be an empty set. Define for message $j = 1, 2$, the conditional total (undetected plus erasure) and undetected error probabilities at terminal $\mathcal{Y}$

$$\xi_j(m_1, m_2) := W_{\mathcal{Y}}^n\big(\mathcal{D}_{m_j}^c \,\big|\, \mathbf{x}(m_1, m_2)\big) \tag{2}$$

$$\hat{\xi}_j(m_1, m_2) := W_{\mathcal{Y}}^n\Big(\bigcup_{\tilde{m}_j \in \mathcal{M}_j \backslash \{m_j\}} \mathcal{D}_{\tilde{m}_j} \,\big|\, \mathbf{x}(m_1, m_2)\Big). \tag{3}$$

Then we may define the average total and undetected error probabilities for message j at terminal $\mathcal{Y}$ as follows:

$$e_j := \frac{1}{|\mathcal{M}_1||\mathcal{M}_2|} \sum_{(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2} \xi_j(m_1, m_2) \tag{4}$$

$$\hat{e}_j := \frac{1}{|\mathcal{M}_1||\mathcal{M}_2|} \sum_{(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2} \hat{\xi}_j(m_1, m_2). \tag{5}$$

The objective of this paper is to find exponential upper bounds for $(e_1, \hat{e}_1, e_2, \hat{e}_2)$, all of which depend on the blocklength n.

III. Decoding Strategy

In this section, we detail the decoding strategy at terminal $\mathcal{Y}$. The decoding strategy and subsequent analysis for terminal $\mathcal{Z}$ is standard and follows from Csiszár and Körner's exposition of decoding with the erasure option [2, Thm. 10.11].

Assume there is a codebook $\mathcal{C}$ consisting *cloud centers* $\mathbf{u}(m_2) \in \mathcal{U}^n, m_2 \in \mathcal{M}_2$ ($\mathcal{U}$ is a finite set) and for each m_2, a set of *satellite codewords* $\mathbf{x}(m_1, m_2) \in \mathcal{X}^n$ indexed by $m_1 \in \mathcal{M}_1$. Fix $\lambda_{12}, \lambda_2 \geq 1$ and $\tilde{R}_j \geq R_j$ for $j = 1, 2$. For brevity, let $\tilde{R}_{12} := \tilde{R}_1 + \tilde{R}_2$ and $R_{12} := R_1 + R_2$ be the sum rates. The decoding rule is given as follows:

Step 1: Decode to $(\hat{m}_1, \hat{m}_2) \in \mathcal{M}_1 \times \mathcal{M}_2$ if and only if this is the *unique* pair of messages such that

$$\begin{aligned} &\hat{I}(\mathbf{u}(\hat{m}_2), \mathbf{x}(\hat{m}_1, \hat{m}_2) \wedge \mathbf{y}) \\ &\quad \geq \tilde{R}_{12} + \lambda_{12}\big|\hat{I}(\mathbf{u}(\tilde{m}_2), \mathbf{x}(\tilde{m}_1, \tilde{m}_2) \wedge \mathbf{y}) - R_{12}\big|^+ \end{aligned} \tag{6}$$

for all $(\tilde{m}_1, \tilde{m}_2) \neq (\hat{m}_1, \hat{m}_2)$. If we cannot find a unique pair of messages satisfying (6), go to Step 2.

Step 2: Declare the first message to be an erasure e_1 and declare the second message to be $\hat{m}_2 \in \mathcal{M}_2$ if and only if it is the *unique* message such that

$$\hat{I}(\mathbf{u}(\hat{m}_2) \wedge \mathbf{y}) \geq \tilde{R}_2 + \lambda_2\big|\hat{I}(\mathbf{u}(\tilde{m}_2) \wedge \mathbf{y}) - R_2\big|^+ \tag{7}$$

for all $\tilde{m}_2 \neq \hat{m}_2$. If we cannot find a unique message satisfying (7), declare the second message to be an erasure e_2 as well.

The intuition behind this two-step algorithm is as follows: In Step 1, we are ambitious. We try to decode both messages M_1 and M_2 using the rule in (6). This rule is a generalization of that for the single-user case in [2, Thm. 10.11]. If decoding *fails* (i.e., no unique message pair satisfies (6)), perhaps due to the stringent choices of $\tilde{R}_1$, $\tilde{R}_2$ and λ_{12}, then we act conservatively. Given $\mathbf{y}$, we *at least* want to decode the cloud center represented by M_2, while we are content with declaring an erasure for M_1. If Step 2 in (7) also fails, we have no choice but to erase both messages. Note that the decoding rules in (6)–(7) are unambiguous because $\lambda_{12}, \lambda_2 \geq 1$ [7, App. I].

IV. Main Result and Interpretation

A. Preliminary Definitions

Before we present the main result, we define a few relevant quantities. First, we fix a joint distribution $P_{UX} \in \mathscr{P}(\mathcal{U} \times \mathcal{X})$. Next fix conditional distributions $V : \mathcal{U} \times \mathcal{X} \to \mathcal{Y}$ and $\hat{V} : \mathcal{U} \to \mathcal{Y}$. Then we may define

$$J_V(R_1, R_2) := |I_V(UX \wedge Y) - R_{12}|^+ \tag{8}$$

$$J_{\hat{V}}(R_2) := |I_{\hat{V}}(U \wedge Y) - R_2|^+. \tag{9}$$

Note that $I_V(UX \wedge Y)$ is the mutual information of UX and Y where the joint distribution of UXY is $P_{UX} \times V$ but P_{UX}, being fixed throughout, is suppressed in the notations in (8) and (9). We define the *marginal and joint modified random coding error exponents for the ABC* as

$$E_{\mathrm{r},\lambda}(\tilde{R}_2) := \min_{\hat{V}} D(\hat{V} \| W_{Y|U} | P_U) + \lambda J_{\hat{V}}(\tilde{R}_2) \tag{10}$$

$$E_{\mathrm{r},\lambda}(\tilde{R}_1, \tilde{R}_2) := \min_{V} D(V \| W | P_{UX}) + \lambda J_V(\tilde{R}_1, \tilde{R}_2). \tag{11}$$

Here, we use the notation $W_{Y|U}$ and W to mean the channels $W_{Y|U}(y|u) := \sum_x W_{\mathcal{Y}}(y|x) P_{X|U}(x|u)$ and $W(y|u, x) := W_{\mathcal{Y}}(y|x)$ for every $u \in \mathcal{U}$. Also note that the exponents in (10) and (11) depend on P_{UX} but this dependence is suppressed. Furthermore, we define the *penalized modified random coding error exponent for the ABC* as

$$\begin{aligned} &E_{\mathrm{r},\lambda}^{-}(\tilde{R}_1, \tilde{R}_2, R_2) \\ &\quad := \min_{V} D(V \| W | P_{UX}) + \lambda J_V(\tilde{R}_1, \tilde{R}_2) - J_V(R_2). \end{aligned} \tag{12}$$

The penalization comes from the fact that we are subtracting the non-negative quantity $J_V(R_2)$ in the optimization above. Define the *sphere packing exponent for the ABC* as

$$E_{\mathrm{sp}}(R) := \min_{V : I_V(UX \wedge Y) \leq R} D(V \| W | P_{UX}). \tag{13}$$

Finally, for $\tilde{R}_2$ and R_2, we define the *difference in rates* $\Delta_2 = \Delta_2(\tilde{R}_2, R_2) := \tilde{R}_2 - R_2$ and similarly, $\Delta_{12} = \Delta_{12}(\tilde{R}_1, R_1, \tilde{R}_2, R_2) := \tilde{R}_{12} - R_{12}$.

B. Main Result

With these preparations, we can now state our main result.

Theorem 1. *There exists a sequence of (n, R_1, R_2)-codes for the ABC such that for any choice of $\tilde{R}_1$, $\tilde{R}_2$, λ_{12}, λ_2 and P_{UX}, we have*

$$\liminf_{n \to \infty} \frac{1}{n} \log \frac{1}{e_1} \geq E_{\mathrm{r},\frac{1}{\lambda_{12}}}^{-}(\tilde{R}_1, \tilde{R}_2, R_2) \tag{14}$$

$$\liminf_{n\to\infty}\frac{1}{n}\log\frac{1}{\hat{e}_1}\ge E^-_{\mathrm{r},\lambda_{12}}(R_1,R_2,R_2)+\Delta_{12} \tag{15}$$

$$\liminf_{n\to\infty}\frac{1}{n}\log\frac{1}{e_2}\ge\min\Big\{E_{\mathrm{r},\lambda_{12}}(R_1,R_2)+\Delta_{12},\ \max\big\{E^-_{\mathrm{r},\frac{1}{\lambda_{12}}}(\tilde{R}_1,\tilde{R}_2,R_2),E_{\mathrm{r},\frac{1}{\lambda_2}}(\tilde{R}_2)\big\}\Big\} \tag{16}$$

$$\liminf_{n\to\infty}\frac{1}{n}\log\frac{1}{\hat{e}_2}\ge\min\Big\{E_{\mathrm{r},\lambda_{12}}(R_1,R_2)+\Delta_{12},\ \max\big\{E^-_{\mathrm{r},\frac{1}{\lambda_{12}}}(\tilde{R}_1,\tilde{R}_2,R_2),E_{\mathrm{r},\lambda_2}(R_2)+\Delta_2\big\}\Big\} \tag{17}$$

C. *Interpretation and Comments*

We now comment on the forms of the exponents.

First, observe from (14) and (15) involve the penalized modified random coding error exponent for the ABC $E^-_{\mathrm{r},\lambda}$, defined in (12). The penalization of the term $J_V(R_2)$ is required because of there are various ways an undetected error and erasure can occur for message 1 according to Step 1 of the decoding rule. More specifically, this penalization results from the kinds of errors that may result from superposition coding [11]: (i) the cloud center may be decoded incorrectly, or (ii) the cloud center is decoded correctly but the satellite is decoded incorrectly. The latter event leads to the term $J_V(R_2)$.

Second, since $\tilde{R}_j\ge R_j$ and $\lambda_{12},\lambda_2\ge 1$, the undetected exponents are at least as large as the error exponents. This is natural since undetected errors are much more undesirable compared to erasures. We can simply re-transmit if there is an erasure. This is known as Automatic Repeat reQuest (ARQ).

Third, observe from (16) and (17) that the exponents associated to message 2 involve several terms. Since the algorithm we proposed in Sec. III involves two distinct steps, undetected errors and erasures for message 2 occur in several different ways so the event with the worst exponent dominates.

Finally, we mention that unlike in Csiszár and Körner [2, Ch. 10] and Körner and Sgarro [9], we do not (explicitly) use the packing lemma (or its generalizations) in our proof. We compute bounds on the error and erasure probabilities averaged over a random constant joint-composition code directly.

V. Proof of Main Result

The proof is split into four subsections, each detailing the calculations leading to the exponents presented in Theorem 1.

Given joint type $P_{UX}\in\mathscr{P}_n(\mathcal{U}\times\mathcal{X})$, we consider the ensemble of codes that are randomly generated as follows: First for each $m_2\in[2^{nR_2}]$, a cloud center $\mathbf{u}(m_2)$ is drawn uniformly at random from the type class $\mathcal{T}_{P_U}\subset\mathcal{U}^n$. Subsequently for each m_2, a cloud of codewords called satellite codewords $\mathbf{x}(m_1,m_2), m_1\in[2^{nR_1}]$ is drawn uniformly at random from the $P_{X|U}$-shell $\mathcal{T}_{P_{X|U}}(\mathbf{u}(m_2))\subset\mathcal{X}^n$. We analyze the error probabilities averaged over this random codebook construction and we assume, without loss of generality, that the transmitted messages are $(M_1,M_2)=(1,1)$.

A. *Undetected Error Probability for Message 1 at Terminal $\mathcal{Y}$*

An undetected error for message 1 occurs if and only if Step 1 succeeds (i.e., outputs a message pair instead of going on to Step 2) but the declared message pair $(\hat{m}_1,\hat{m}_2)$ is such that the first component is some natural number other than 1 (second component is arbitrary). In other words, we have

$$\hat{e}_1=\Pr\Bigg(\bigcup_{\tilde{m}_1\in\mathcal{M}_1\setminus\{1\},\tilde{m}_2\in\mathcal{M}_2}\mathcal{E}(\tilde{m}_1,\tilde{m}_2)\Bigg) \tag{18}$$

where the event $\mathcal{E}(\tilde{m}_1,\tilde{m}_2)$ is defined as

$$\mathcal{E}(\tilde{m}_1,\tilde{m}_2):=\big\{\hat{I}(U^n(\tilde{m}_2),X^n(\tilde{m}_1,\tilde{m}_2)\wedge Y^n)\ge\tilde{R}_{12}+\lambda_{12}\big|\hat{I}(U^n(1),X^n(1,1)\wedge Y^n)-R_{12}\big|^+\big\} \tag{19}$$

To analyze $\hat{e}_1$, we first condition on $(U^n(1),X^n(1,1),Y^n)$ having various joint types, i.e.,

$$\hat{e}_1:=\sum_{V_{UXY}}\sum_{(\mathbf{u},\mathbf{x},\mathbf{y})\in\mathcal{T}_{V_{UXY}}}P_{U^nX^n}(\mathbf{u},\mathbf{x})W^n_{\mathcal{Y}}(\mathbf{y}|\mathbf{x})\times\Pr\Bigg(\bigcup_{\tilde{m}_1\in\mathcal{M}_1\setminus\{1\},\tilde{m}_2\in\mathcal{M}_2}\mathcal{E}(\tilde{m}_1,\tilde{m}_2)\,\Bigg|\,\mathbf{u},\mathbf{x},\mathbf{y}\Bigg) \tag{20}$$

The first sum over V_{UXY} is in fact over all joint types in $\mathscr{P}_n(\mathcal{U}\times\mathcal{X}\times\mathcal{Y})$ for which the $(\mathcal{U}\times\mathcal{X})$-marginal is P_{UX}. The conditioning in the probability in (20) is on the event $\{U^n(1)=\mathbf{u},X^n(1,1)=\mathbf{x},Y^n=\mathbf{y}\}$ but we shorten this to $\{\mathbf{u},\mathbf{x},\mathbf{y}\}$ for brevity.

Now we distinguish between two cases: Case (i) $\tilde{m}_1\ne 1,\tilde{m}_2\ne 1$, and Case (ii) $\tilde{m}_1\ne 1,\tilde{m}_2=1$. For Case (i), there are $\lceil 2^{nR_{12}}\rceil-1$ such events and by symmetry we may analyze

$$\Pr\big(\mathcal{E}(2,2)\,\big|\,\mathbf{u},\mathbf{x},\mathbf{y}\big)=\Pr\big(\hat{I}(U^n(2),X^n(2,2)\wedge Y^n)\ge t\,\big|\,\mathbf{u},\mathbf{x},\mathbf{y}\big) \tag{21}$$

where given $(\mathbf{u},\mathbf{x},\mathbf{y})$, the parameter

$$t=\tilde{R}_{12}+\lambda_{12}\,|I_V(UX\wedge Y)-R_{12}|^+ \tag{22}$$

is fixed. We suppress the dependence of t on $(\mathbf{u},\mathbf{x},\mathbf{y})$. Now we bound the probability in (21) as follows:

$$\Pr\big(\mathcal{E}(2,2)\,\big|\,\mathbf{u},\mathbf{x},\mathbf{y}\big)=\sum_{\substack{\hat{V}_{UX|Y}\in\mathscr{V}_n(\mathcal{U}\times\mathcal{X};P_{\mathbf{y}}):\\ P_{\mathbf{y}}\hat{V}_{UX|Y}=P_{UX},I(P_{\mathbf{y}},\hat{V}_{UX|Y})\ge t}}P_{U^nX^n}(\mathcal{T}_{\hat{V}_{UX|Y}}(\mathbf{y})) \tag{23}$$

$$\doteq\sum_{\substack{\hat{V}_{UX|Y}\in\mathscr{V}_n(\mathcal{U}\times\mathcal{X};P_{\mathbf{y}}):\\ P_{\mathbf{y}}\hat{V}_{UX|Y}=P_{UX},I(P_{\mathbf{y}},\hat{V}_{UX|Y})\ge t}}\exp(-nI(P_{\mathbf{y}},\hat{V}_{UX|Y})) \tag{24}$$

$$\doteq\exp(-nt) \tag{25}$$

where (24) follows from a standard method of types calculation. See, for example, [2, Lem 10.1] or [12, Appendix]. Next consider Case (ii). In this case there are at most 2^{nR_1} such events which indicates that the cloud center is decoded correctly but the satellite codeword is not. The conditional probability of a generic event in this case $\mathcal{E}(2,1)$ can be bounded as follows:

$$\Pr\big(\mathcal{E}(2,1)\,\big|\,\mathbf{u},\mathbf{x},\mathbf{y}\big)=\sum_{\substack{\hat{V}_{X|UY}\in\mathscr{V}_n(\mathcal{X};P_{\mathbf{u},\mathbf{y}}):\\ P_{\mathbf{y}|\mathbf{u}}\hat{V}_{X|UY}=P_{X|U},\\ I(P_{\mathbf{u},\mathbf{y}},\hat{V}_{X|UY})\ge t}}P_{X^n|U^n}(\mathcal{T}_{\hat{V}_{X|UY}}(\mathbf{u},\mathbf{y})|\mathbf{u}) \tag{26}$$

$$\doteq \sum_{\substack{\hat{V}_{X|UY}\in\mathscr{V}_n(\mathcal{X};P_{\mathbf{u},\mathbf{y}}):\\ P_{\mathbf{y}|\mathbf{u}}\hat{V}_{X|UY}=P_{X|U},\\ I(P_{\mathbf{y}|\mathbf{u}},\hat{V}_{X|UY}|P_{\mathbf{u}})+I(P_{\mathbf{u}},P_{\mathbf{y}|\mathbf{u}})\ge t}} \exp(-nI(P_{\mathbf{y}|\mathbf{u}},\hat{V}_{X|UY}|P_{\mathbf{u}})) \tag{27}$$

$$\doteq \exp(-n[t - I_V(U\wedge Y)]) \tag{28}$$

where in the last step, we recall that the joint type of $(U^n(1), X^n(1,1), Y^n)$ is V_{UXY} so we have the equality $I_V(U\wedge Y) = I(P_{\mathbf{u}}, P_{\mathbf{y}|\mathbf{u}})$. Note that the exponent here is $t - I_V(U\wedge Y)$, which is different from the exponent resulting from the calculation leading to (25), i.e., simply t. Putting the bounds in (25) and (28) together and applying the union bound to the probability in (20), we obtain

$$\Pr\Big(\bigcup_{\tilde{m}_1\neq 1,\tilde{m}_2}\mathcal{E}(\tilde{m}_1,\tilde{m}_2)\,\Big|\,\mathbf{u},\mathbf{x},\mathbf{y}\Big)$$

$$\le 2^{nR_{12}}\Pr\big(\mathcal{E}(2,2)\,\big|\,\mathbf{u},\mathbf{x},\mathbf{y}\big) + 2^{nR_1}\Pr\big(\mathcal{E}(2,1)\,\big|\,\mathbf{u},\mathbf{x},\mathbf{y}\big) \tag{29}$$

$$\doteq \exp(-n\min\{t-R_{12}, t-R_1-I_V(U\wedge Y)\}) \tag{30}$$

$$= \exp\big(-n\big(t - R_{12} - \big|I_V(U\wedge Y) - R_2\big|^+\big)\big). \tag{31}$$

Since $P_{U^nX^n}(\mathbf{u},\mathbf{x}) \mathrel{\dot{\le}} P^n_{UX}(\mathbf{u},\mathbf{x})$ for every $(\mathbf{u},\mathbf{x})$,

$$\sum_{(\mathbf{u},\mathbf{x},\mathbf{y})\in\mathcal{T}_{V_{UXY}}} P_{U^nX^n}(\mathbf{u},\mathbf{x})W^n_{\mathcal{Y}}(\mathbf{y}|\mathbf{x}) \mathrel{\dot{\le}} \exp(-nD(V\|W|P_{UX})). \tag{32}$$

Since there are only polynomially many joint types [2, Ch. 2], uniting (22), (31) and (32), we conclude that the undetected error exponent for decoding message 1 is

$$\liminf_{n\to\infty}\frac{1}{n}\log\frac{1}{\hat{e}_1}$$

$$\ge \min_V D(V\|W|P_{UX}) + \lambda_{12}\,|I_V(UX\wedge Y) - R_{12}|^+$$

$$- |I_V(U\wedge Y) - R_2|^+ + \Delta_{12} \tag{33}$$

$$= E^-_{\mathrm{r},\lambda_{12}}(R_1,R_2,R_2) + \Delta_{12}, \tag{34}$$

where the last line arises from the definition of the penalized modified random coding exponent in (12). This gives (15).

B. *Total Error Probability for Message 1 at Terminal $\mathcal{Y}$*

An error for message 1 occurs under two possible conditions: (i) Step 1 succeeds, in which case we have an undetected error since m_1 is declared to be some natural number not equal to 1; (ii) Step 1 fails, in which case m_1 is erased. We already analyzed Case (i) in Sec. V-A and because $\tilde{R}_j \ge R_j, j=1,2$ and $\lambda_{12}\ge 1$, this case will not dominate, i.e., its exponent will be larger than that for Case (ii). Hence, we focus on Case (ii), i.e., there is not unique message pair that satisfies (6). In particular, message pair $(1,1)$ does *not* satisfy (6). Thus,

$$e_1 \doteq \Pr\Big(\bigcup_{(\tilde{m}_1,\tilde{m}_2)\in(\mathcal{M}_1\setminus\{1\})\times(\mathcal{M}_2\setminus\{1\})}\mathcal{J}(\tilde{m}_1,\tilde{m}_2)\Big) \tag{35}$$

where the event $\mathcal{J}(\tilde{m}_1,\tilde{m}_2)$ is defined as follows:

$$\mathcal{J}(\tilde{m}_1,\tilde{m}_2) := \big\{\hat{I}(U^n(1), X^n(1,1)\wedge Y^n)$$

$$\le \tilde{R}_{12} + \lambda_{12}\big|\hat{I}(U^n(\tilde{m}_2), X^n(\tilde{m}_1,\tilde{m}_2)\wedge Y^n) - R_{12}\big|^+\big\}. \tag{36}$$

Now similarly to (20), we again partition into various joint types, i.e.,

$$e_1 \doteq \sum_{V_{UXY}}\sum_{(\mathbf{u},\mathbf{x},\mathbf{y})\in\mathcal{T}_{V_{UXY}}} P_{U^nX^n}(\mathbf{u},\mathbf{x})W^n_{\mathcal{Y}}(\mathbf{y}|\mathbf{x})$$

$$\times\Pr\Big(\bigcup_{(\tilde{m}_1,\tilde{m}_2)\in(\mathcal{M}_1\setminus\{1\})\times(\mathcal{M}_2\setminus\{1\})}\mathcal{J}(\tilde{m}_1,\tilde{m}_2)\,\Big|\,\mathbf{u},\mathbf{x},\mathbf{y}\Big) \tag{37}$$

$$=: e_{1,\mathcal{A}} + e_{1,\mathcal{A}^c}. \tag{38}$$

In (38), we split the analysis into two parts by partitioning the joint types $V_{UXY} = P_{UX}\times V \in \mathscr{P}_n(\mathcal{U}\times\mathcal{X}\times\mathcal{Y})$ into two classes: $\mathcal{A} := \{V_{UXY} : I(P_{UX},V)\le\tilde{R}_{12}\}$ and $\mathcal{A}^c := \{V_{UXY} : I(P_{UX},V) > \tilde{R}_{12}\}$. The first class results in a sphere packing-like bound. More precisely, by Sanov's theorem [2, Prob. 2.12],

$$e_{1,\mathcal{A}} = \sum_{V_{UXY}\in\mathcal{A}}\sum_{(\mathbf{u},\mathbf{x},\mathbf{y})\in\mathcal{T}_{V_{UXY}}} P_{U^nX^n}(\mathbf{u},\mathbf{x})W^n_{\mathcal{Y}}(\mathbf{y}|\mathbf{x}) \tag{39}$$

$$\doteq \exp(-nE_{\mathrm{sp}}(\tilde{R}_{12})). \tag{40}$$

In the last line, we employed the definitions of $\mathcal{A}$ and that of the sphere packing exponent in (13).

Now we deal with the other joint types, i.e., those in $\mathcal{A}^c$. Again, we partition the analysis into three cases: Case (i) $\tilde{m}_1\neq 1, \tilde{m}_2\neq 1$; Case (ii) $\tilde{m}_1 = 1, \tilde{m}_2\neq 1$; Case (iii) $\tilde{m}_1 = 1, \tilde{m}_2 \neq 1$.

For Case (i), there are a total of $2^{nR_{12}}$ events with identical probability. A generic such probability is

$$\Pr\big(\mathcal{J}(2,2)\,\big|\,\mathbf{u},\mathbf{x},\mathbf{y}\big)$$

$$\le \Pr\big(\hat{I}(U^n(2), X^n(2,2)\wedge Y^n)\ge s\,\big|\,\mathbf{u},\mathbf{x},\mathbf{y}\big) \tag{41}$$

where now because the joint type V_{UXY} is in $\mathcal{A}^c$, $\hat{I}(U^n(1), X^n(1,1)\wedge Y^n) > \tilde{R}_{12}$ (i.e., the $|\cdot|^+$ is inactive), the parameter s in (41) is defined as

$$s := \frac{1}{\lambda_{12}}|I_V(UX\wedge Y) - \tilde{R}_{12}|^+ + R_{12}. \tag{42}$$

By the same reasoning as the steps leading to (25), we have

$$\Pr\big(\mathcal{J}(2,2)\,\big|\,\mathbf{u},\mathbf{x},\mathbf{y}\big) \mathrel{\dot{\le}} \exp(-ns). \tag{43}$$

For Case (ii), there are a total of 2^{nR_1} events with identical probability. Similarly to the calculation that led to (28) and (43), we have that

$$\Pr\big(\mathcal{J}(2,1)\,\big|\,\mathbf{u},\mathbf{x},\mathbf{y}\big) \mathrel{\dot{\le}} \exp(-n[s - I_V(U\wedge Y)]) \tag{44}$$

For Case (iii), there are a total of 2^{nR_2} events with identical probability. Note that here the cloud center, represented by $\tilde{m}_2$, is incorrect, so similarly to the calculation that led to (43), we have that

$$\Pr\big(\mathcal{J}(1,2)\,\big|\,\mathbf{u},\mathbf{x},\mathbf{y}\big) \mathrel{\dot{\le}} \exp(-ns). \tag{45}$$

Case (iii), however, will not dominate since there are only 2^{nR_2} events each with probability given by (45). This is dominated by Case (i) in which there are exponentially many more error events ($2^{nR_{12}}$ to be precise) with the same bound on

the error probability. So we may safely omit the contribution by Case (iii).

Putting the analysis for the joint types $V_{UXY} \in \mathcal{A}^c$ together and using (32), we obtain that

$$\liminf_{n\to\infty} \frac{1}{n} \log \frac{1}{e_{1,\mathcal{A}^c}} \geq E^-_{\mathrm{r},\frac{1}{\lambda_{12}}}(\tilde{R}_1, \tilde{R}_2, R_2) \tag{46}$$

By using the fact that $|\cdot|^+ \geq 0$,

$$E^-_{\mathrm{r},\frac{1}{\lambda_{12}}}(\tilde{R}_1, \tilde{R}_2, R_2) \leq E_{\mathrm{r},\frac{1}{\lambda_{12}}}(\tilde{R}_1, \tilde{R}_2). \tag{47}$$

By using weak duality in optimization theory, it can be seen that

$$E_{\mathrm{r},\frac{1}{\lambda_{12}}}(\tilde{R}_1, \tilde{R}_2) \leq E_{\mathrm{sp}}(\tilde{R}_{12}). \tag{48}$$

Hence, contribution from the joint types in $\mathcal{A}$ given by the calculation in (40) do not dominate. As a result, the error exponent for decoding message 1 is dominated by the joint types in $\mathcal{A}^c$, and so the exponential behaviors of the upper bounds of e_1 and $e_{1,\mathcal{A}^c}$ are the same. We thus conclude that

$$\liminf_{n\to\infty} \frac{1}{n} \log \frac{1}{e_1} \geq E^-_{\mathrm{r},\frac{1}{\lambda_{12}}}(\tilde{R}_1, \tilde{R}_2, R_2). \tag{49}$$

C. Undetected Error Probability for Message 2 at Terminal $\mathcal{Y}$

An undetected error for message 2 occurs in one of two ways: (i) Step 1 succeeds but the declared message pair $(\hat{m}_1, \hat{m}_2)$ is such that the second component is some natural number not equal to 1; (ii) Step 1 fails (which we denote as event $\mathcal{F}$) and Step 2 succeeds but the declared message in this step $\hat{m}_2$ is some natural number not equal to 1. Thus,

$$\hat{e}_2 = \Pr\Bigg(\bigcup_{\tilde{m}_1 \in \mathcal{M}_1 \tilde{m}_2 \in \mathcal{M}_2 \setminus \{1\}} \mathcal{E}(\tilde{m}_1, \tilde{m}_2) \\ \cup \Big(\mathcal{F} \cap \bigcup_{\tilde{m}_2 \in \mathcal{M}_2 \setminus \{1\}} \mathcal{G}(\tilde{m}_2) \Big) \Bigg) \tag{50}$$

where $\mathcal{E}(\tilde{m}_1, \tilde{m}_2)$ is defined in (19) and

$$\mathcal{G}(\tilde{m}_2) := \big\{ \hat{I}(U^n(\tilde{m}_2) \wedge Y^n) \\ \geq \tilde{R}_2 + \lambda_2 \big| \hat{I}(U^n(1) \wedge Y^n) \geq R_2 \big|^+ \big\}. \tag{51}$$

For Case (i), by using a similar calculation to that in Sec. V-A, we have

$$\liminf_{n\to\infty} \frac{1}{n} \log \frac{1}{\Pr\left(\bigcup_{\tilde{m}_1 \in \mathcal{M}_1 \tilde{m}_2 \in \mathcal{M}_2 \setminus \{1\}} \mathcal{E}(\tilde{m}_1, \tilde{m}_2) \right)} \\ \geq E_{\mathrm{r},\lambda_{12}}(R_1, R_2) + \Delta_{12}. \tag{52}$$

An important point to note here is that the term $J_V(R_2)$ is *absent* because here we do not have to bound the probability that the cloud center is decoded correctly but the satellite codeword is decoded incorrectly. So the exponent here is the *unpenalized* random coding error exponent for the ABC.

Now for Case (ii), we first analyze the probability that Step 1 fails, i.e., there is no unique $(\hat{m}_1, \hat{m}_2)$ satisfying (6). This exponent is exactly that calculated in Sec. V-B. Thus,

$$\liminf_{n\to\infty} \frac{1}{n} \log \frac{1}{\Pr(\mathcal{F})} \geq E^-_{\mathrm{r},\frac{1}{\lambda_{12}}}(\tilde{R}_1, \tilde{R}_2, R_2). \tag{53}$$

Note that the *penalization is* present here because $(\tilde{m}_1, \tilde{m}_2) \neq (1,1)$ means there are three cases: (a) $\tilde{m}_1 \neq 1, \tilde{m}_2 \neq 1$; (b) $\tilde{m}_1 = 1, \tilde{m}_2 \neq 1$; and (c) $\tilde{m}_1 \neq 1, \tilde{m}_2 = 1$. Finally, we need to bound the probability that the declared message in Step 2 is some natural number not equal to 1. This yields

$$\liminf_{n\to\infty} \frac{1}{n} \log \frac{1}{\Pr\left(\bigcup_{\tilde{m}_2 \in \mathcal{M}_2 \setminus \{1\}} \mathcal{G}(\tilde{m}_2) \right)} \geq E_{\mathrm{r},\lambda_2}(R_2) + \Delta_2. \tag{54}$$

So the exponent for Case (ii) is the maximum of the exponents derived in (53) and (54). Combining all these exponents yields the undetected error probability for message 2 in (17).

D. Total Error Probability for Message 2 at Terminal $\mathcal{Y}$

Finally, we compute the total error probability for message 2. An error occurs if and only if one of two events occurs: (i) Step 1 succeeds but message 2 is declared to be some $\hat{m}_2 \in \mathcal{M}_1 \setminus \{1\}$ (i.e., undetected error) or (ii) Step 1 fails and an error (undetected or erasure) occurs in Step 2.

For Case (i), the exponent of the error probability is $E_{\mathrm{r},\lambda_{12}}(\tilde{R}_1, \tilde{R}_2) + \Delta_{12}$ without penalization because the cloud center $\tilde{m}_2$ suffers from an undetected error.

In Case (ii), Step 1 fails (event $\mathcal{F}$ in Sec. V-C) and an error occurs in Step 2. Step 1 failing results in an error exponent of $E^-_{\mathrm{r},\frac{1}{\lambda_{12}}}(\tilde{R}_1, \tilde{R}_2, R_2)$; cf. (53). An error occurs in Step 2 with exponent $E_{\mathrm{r},\frac{1}{\lambda_2}}(\tilde{R}_2)$ by the same reasoning as the calculations in Sec. V-B. Combining these exponents yields (16).

Acknowledgements

Discussions with Pierre Moulin and Silas L. Fong are gratefully acknowledged. The author's research is supported by an NUS Young Investigator Award R-263-000-B37-133.

References

[1] G. D. Forney, "Exponential error bounds for erasure, list, and decision feedback schemes," *IEEE Trans. on Inf. Th.*, vol. 14, no. 2, pp. 206–220, 1968.

[2] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.

[3] E. Telatar, "Multi-access communications with decision feedback decoding," Ph.D. dissertation, Massachusetts Institute of Technology, 1992.

[4] P. Moulin, "A Neyman-Pearson approach to universal erasure and list decoding," *IEEE Trans. on Inf. Th.*, vol. 55, no. 10, pp. 4462–4478, 2009.

[5] N. Merhav, "Error exponents of erasure/list decoding revisited via moments of distance enumerators," *IEEE Trans. on Inf. Th.*, vol. 54, no. 10, pp. 4439–4447, 2008.

[6] A. Somekh-Baruch and N. Merhav, "Exact random coding exponents for erasure decoding," *IEEE Trans. on Inf. Th.*, vol. 57, no. 10, pp. 6444–6454, 2011.

[7] E. Sabbag and N. Merhav, "Achievable error exponents for channels with side information–erasure and list decoding," *IEEE Trans. on Inf. Th.*, vol. 56, no. 11, pp. 5424–5431, 2010.

[8] J. Körner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. on Inf. Th.*, vol. 23, no. 1, pp. 60–64, 1977.

[9] J. Körner and A. Sgarro, "Universally attainable error exponents for broadcast channels with degraded message sets," *IEEE Trans. on Inf. Th.*, vol. 26, no. 6, pp. 670–679, 1980.

[10] Y. Kaspi and N. Merhav, "Error exponents for broadcast channels with degraded message sets," *IEEE Trans. on Inf. Th.*, vol. 57, no. 1, pp. 101–123, 2010.

[11] T. Cover, "Broadcast channels," *IEEE Trans. on Inf. Th.*, vol. 18, no. 1, pp. 2–14, 1972.

[12] V. Y. F. Tan, "On the reliability function of the discrete memoryless relay channel," *IEEE Trans. on Inf. Th.*, vol. 61, 2015.